



به نام هستی بخش

"و سلام بر مهدی که انتظارش را نه فقط دل عاشق،
که ترنم هر باران بهاری و هر روزنهی امید میکشد..."

تمرین 8 (فصل 12 و 13)

موعد تحویل:

درس پایگاه داده ها، بهار 1403

1. به سوالات زیر پاسخ دهید.

الف) استفاده از متد Chunked-IO چگونه هزینه IO را کاهش میدهد؟ چه trade-off ای بین block-size و هزینه وجود دارد؟

ب) کدام بخش از عملگر projection هزینه بر است؟ روش های آن را نام برده و توضیح کوتاه دهید.

ج) آیا میتوان در یک پلن fully-piped-line از روش sort-merge-join استفاده کرد؟ توضیح دهید.

د) سناریویی مثال بزنید که هزینه external merge sort با internal merge sort یکی باشد. (N و B را پیدا کنید).

2. برای دسترسی به هر یک از رکورد های زیر از کدام شاخص hash یا b+ tree بر روی (a,b,c,d) میتوان استفاده کرد؟ دلیل انتخاب یا عدم انتخاب هر شاخص را بیان کنید.

a) $a > 9 \text{ AND } b < 8$

b) $a = 1 \text{ AND } (b < 7 \text{ OR } c = 2)$

c) $a = 1 \text{ OR } (b < 7 \text{ AND } c = 2)$

3. فرض کنید query روبه رو به شما داده شده : $\text{day} < 8/9/98 \text{ AND } \text{bid} = 6 \text{ AND } \text{sid} = 102$ و شاخص های b+ tree بر روی day و hash روی (bid,sid) است.

الف) با تعیین کردن نوع clustered/unclustered بودن هر کدام، 3 سناریو طراحی کنید که در اولی شاخص hash بهتر باشد، در دومی شاخص b+ tree و در سومی scan کردن کل فایل ها بهتر باشد.

(فرض کنید کل relation دارای 100 صفحه و 20000 رکورد است و reduction factor هر ترم 0.1 است)

ب) query optimizer از کجا متوجه میشود که کدام سناریو پیش آمده و اطلاعات relation را از کجا دارد؟

4. با توجه به شمای زیر و توضیحات آن به سوالات زیر پاسخ دهید:

Students (sid, sname, gpa): 300 Pages, 20 tuple/page

Enrollments (sid, cid, quantity): 600 pages, 150 tuple/page

Courses (cid, cname, classNo): 150 pages, 20 tuple/page

یک شاخص clustered tree روی sid در students

یک شاخص unclustered hash روی sid بر روی enrollments

فرض کنید هر دانشجو به طور متوسط 10 کورس دارد.

الف) کمترین هزینه join کردن دو رابطه students و enrollments را به ازای دو روش INLJ و PONLJ به دست بیاورید.

ب) اگر قرار باشد از روش external merge sort join استفاده کنیم، فقط برای بخش sort کردن، هزینه چقدر میشود؟ (تعداد صفحه های بافر را 5 در نظر بگیرید)

5. این سوال را نیز با توجه شمای سوال قبل و شاخص های زیر:

شاخص های students:

unclustered hash on sid

شاخص های courses:

unclustered hash on sid

unclustered tree on cname

شاخص های enrollments:

unclustered tree on quantity

unclustered tree on cid

برای query زیر، پلن با کمترین هزینه را نشان داده و هزینه آن را به دست آورید:

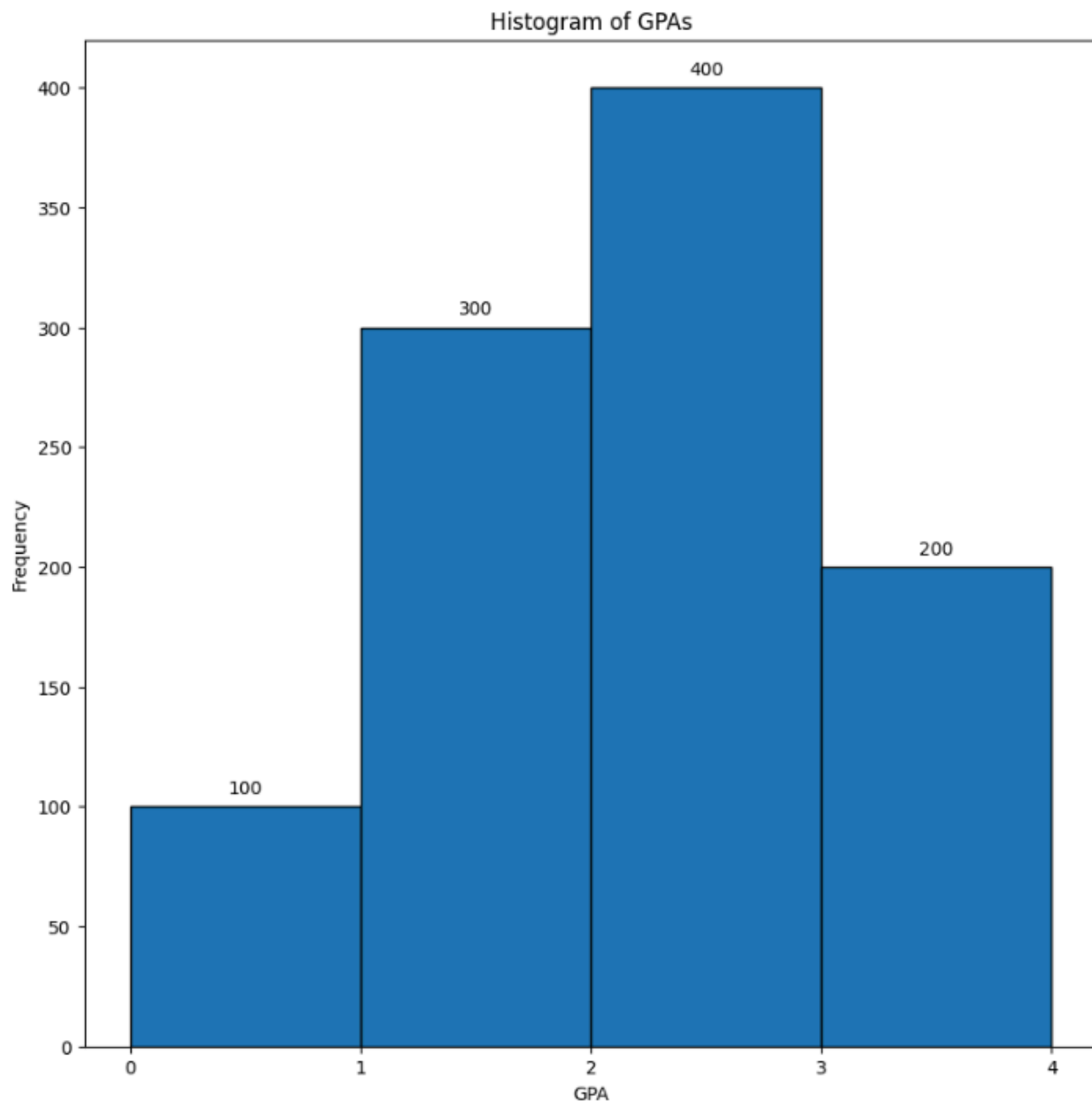
Select S.sid, C.cid

From Students S, Enrollments E, Courses C

Where S.sid = E.sid AND C.cid = E.cid AND S.gpa >= 2 AND E.quantity > 40 AND C.cname = 'Database'

فرض کنید هر دانشجو به طور متوسط 10 کورس دارد.

فرض کنید که توزیع quantity کلاس ها به صورت یکنواخت از 10 تا 60 است، برای درس دیتابیس 20 کورس برگزار شده و توزیع gpa دانشجویان به شکل زیر است:



6. اگر بخواهیم فایلی با 10000 صفحه را با 8 صفحه بافر و با روش internal merge sort مرتب کنیم:

الف) چه تعداد pass برای مرتب سازی نیاز است؟ هزینه IO چقدر خواهد شد؟

ب) تعداد و اندازه run های تولید شده در هر pass را مشخص کنید

ج) فرض کنید می‌خواهیم هزینه IO را با روش chunked-IO کاهش دهیم، سائز بلاک را طوری به دست بیاورید که تعداد pass ها ثابت بماند.

سربلند و پیروز باشید